DOCUMENT RESUME

ED 298 154 TM 012 250

AUTHOR Wallace, Gaylen R.
TITLE RSE-40: An Alternate Scoring System for the Rosenberg Self-Esteem Scale (RSE).
PUB DATE Jan 88
NOTE 17p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adults; College Students; Higher Education; High Schools; High School Students; *Likert Scales; *Psychometrics; Scoring; *Scoring Formulas; *Self Esteem; *Test Reliability; Test Validity
IDENTIFIERS Guttman Scales; *Rosenberg Self Esteem Scale

ABSTRACT

The Rosenberg Self-Esteem Inventory (RSE) is a 10-item scale purporting to measure self-esteem using self-acceptance and self-worth statements. This analysis covers concerns about the degree to which the RSE items represent a particular content universe, the RSE's applicability, factor analytic methods used, and the RSE's reliability and validity. A new Likert scoring method is proposed as an alternative to that used for the original scale (the Guttman scaling method). Besides being difficult to score, criticisms of the Guttman scaling technique exist in the psychometric literature. The newly proposed scoring method, unlike the previously used method, seems to be able to distinguish among several personality and task-specific variables. Both the Guttman scaled RSE and the Likert scaled RSE appear to be unidimensional and seem to have adequate reliability and validity. However, since the response continuum of the original RSE has all the characteristics of the Likert scaling method and the 40-point Likert method has been used in other studies, use of the Likert scoring system is recommended. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED298154

RSE-40: AN ALTERNATE SCORING SYSTEM FOR THE ROSENBERG SELF-ESTEEM SCALE (RSE)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

GAYLEN R. WALLACE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Gaylen R. Wallace, Ed.D.
Associate Professor of Education
Department of Behavioral Studies
University of Missouri-St. Louis
St. Louis, MO 63121

Unpublished Manuscript, Department of Behavioral Studies, University of Missouri-St. Louis, St. Louis, MO 63121. January, 1988.

TM 012 250

BEST COPY AVAILABLE

RSE-40: AN ALTERNATIVE SCORING SYSTEM FOR THE ROSENBERG SELF-ESTEEM SCALE (RSE)

Gaylen R. Wallace, Ed.D.

RSE Original Scoring System

The Rosenberg Self-Esteem inventory (RSE) is a 10 item scale purporting to measure self-esteem using self-acceptance and self-worth statements (Rosenberg, 1965, 1979). The RSE was developed using the standard Guttman scaling technique. The system Rosenberg used to score the original RSE was a "contrived item" scaling procedure suggested by Stouffer et al., (1952). These "contrived" Scale Items and their related RSE items (see Appendix A) are summarized here for clarity.

Contrived Scale Items	RSE Items
Scale Item I	RSE items 3,7,9.
Scale Item II	RSE items 4,5.
Scale Item III	RSE item 1.
Scale Item IV	RSE item 8.
Scale Item V	RSE item 10.
Scale Item VI	RSE items 2,6.

The "contrived item" scoring system yields a seven point (0 - 6) total scale score for each individual. Each Scale Item is assigned a score of either 1 or 0 based upon how one responds to the individual RSE items associated with each "contrived" scale item. Scale Item I is "contrived" from (made up of) RSE items 3,7 and 9. If an individual answers any two of three or three of three statements in the keyed direction they are assigned a score of 1. If, on the other hand, they answer less than two statements in the

keyed direction a score of zero is assigned. For example, by responding as SA (Strongly Agree) or A (Agree) on RSE items 3 and 7 and D (Disagree) or SD (Strongly Disagree) on item 9 an individual would have three of three appropriately keyed responses and would receive a Scale Item score of 1. On the other hand, be responding

---

Scale Item I

RSE Item 3. I feel that I have a number of good qualities.

SA* A* D SD

RSE Item 7. I feel that I'm a person of worth, at least on an equal plane with others.

SA* A* D SD

RSE Item 9. All in all, I am inclined to feel that I am a failure.

SA A D* SD*

* Represents Keyed Response

---

with a D or SD for RSE items 3 and 7 and SA or A to item 9 an individual would have three "nonkeyed" responses and would receive a Scale Item score of 0. On Scale Items II and VI, a Scale Item score of 1 is assigned if an individual gets one or both of the items in the keyed direction and a Scale Item score of 0 if neither of the items are scored in the keyed direction. Scale Items III, IV, and V are single items and would be scored as 1 in the keyed direction and as 0 if missed. An individual's total score is obtained by summing the scores from these six Scale Items and would range from zero to six (where zero is a score).

Concerns

Unidimensionality

One of the major psychometric concerns about any multi-item scale is the degree to which the items represent a particular content universe. Does the scale tap one dimension or domain? One way of creating a unidimensional scale is to select the "best" group of items which will in turn differentiate between persons who have the characteristic in question from those who do not possess that characteristic. Rosenberg (1965, 1979) states that the RSE was developed as a unidimensional scale. Others have also treated the RSE as a homogeneous or unidimensional scale (Fisher, 1972; Lee, 1972; Lieberman, Yalom & Miles, 1972; Marsland & Perry, 1972; Miller, 1973; Stang, 1972; Tessler & Schwartz, 1972; Williams & Stockman, 1973).

Guttman Technique. Rosenberg (1965) employed the Guttman scaling technique to help select his sample of items for the original RSE scale. He choose only those items that dealt with a favorable or unfavorable attitude toward the self by using the Guttman technique (Nunnally, 1974; Torgerson, 1958). This technique requires that increasing proportions of subjects endorse each succeeding item as being true of themselves. Each successive item represents differing degrees of strength for the construct in question, in this case self-esteem. The items which met this criteria were then grouped together as a unit (scale) purporting to measure the same construct. Wylie (1974) comments that "Rosenberg (1965) is apparently the only person who has tried to achieve a unidimensional measure of global self-regard (called by him "self-

esteem") through the method of Guttman scaling" (p. 181).

In spite of its intuitive appeal there are several concerns about the RSE scale and the Guttman technique. First, although the RSE claims to be a Guttman type scale, Crandall (1973) states that in the development of the items which make up the RSE scale Rosenberg does not discuss how the scale was developed, except that it was meant to be a Guttman scale. Nor does he discuss the larger pool of items from which the 10 items making up the RSE were drawn. In addition, the Guttman technique is often criticized. Nunnally (1967) summarizes the criticism by stating:

> In summary, the deterministic model underlying the Guttman scale is thoroughly illogical for most psychological measurement because (1) almost no items exist that fit the model, (2) the presence of a triangular pattern is a necessary but not sufficient condition for the fit of the model in particular instances, (3) the triangular pattern can be (and usually is) artificially forced by dealing with a small number of items that vary greatly in difficulty, (4) the model aspires only to develop ordinal scales, and (5) there are better intuitive bases for developing models for psychological attributes. Considering this heavy weight of criticism, it is surprising that some people still consider this deterministic model a good basis for developing measures of psychological attributes. (p. 66).

Factor Analysis. Another way of demonstrating unidimensionality is to apply a factor analysis statistical procedure. This technique is often used to "discover" clusters of "like items" or factors. A number of factor analytic studies, which use the Likert scoring procedure instead of the Guttman scoring technique, are reported in the research literature (Carmines & Zeller, 1974; Hensley & Roberts, 1976; Kaplan (1970); Kaplan & Pokorney, 1969, 1971; Kohn, 1969; Zeller & Carmines, 1976). In particular Kaplan & Pokorney (1969, 1971) obtained two uncorrelated factors which accounted for 45 percent of the total score variance. They

concluded that "the 10 items employed by Rosenberg were multifactorial in nature" (p. 425). On the other hand Carmines & Zeller (1974) report that their two factors seemed "to tap the same rather than different dimensions, for their correlations with the criterion variables are almost identical to one another in terms of direction, strength, and consistency" (p.). Hensley & Roberts (1976) found a two factor solution using college students but found that one factor represented positive items and one represented negative items. Although these studies may seem to support two separate factors, Hensley & Roberts (1976) reflect what is probably the more accurate perspective when they state, "We conclude the factors identify an underlying response set and that the scale probably measures a single variable" (p. 583). Therefore, until more conclusive evidence is presented, Rosenberg's (1965, 1979) assertions that the RSE is a homogeneous or unidimensional scale is the safer conclusion.

Reliability and Validity

The "contrived score" based RSE, though awkward in it's scoring procedure, seems to have adequate reliability and validity. Reliability coefficients of reproducibility have been reported to range form .91 to .93 and scalability to range from .71 to .73 (Kahle, 1976; Nocks & Bradley, 1969; Rosenberg, 1965; Wylie, 1974). Silbert & Tippett (1965) report a test-retest coefficient of .85. Construct validity coefficients with other self-report scales ranged from .21 to .83 and criterion validity coefficients with other variables ranged from -.38 to .83 (Crandall, 1973; Friedel, et al., 1979; Nocks & Bradley, 1969; Silbert & Tippett, 1965; Yamamota &

Wiersma, 1967).

In general then, the Guttman scaled RSE appears to be unidimensional and homogeneous and the items seem to represent a "best set" from a larger item pool. These items apparently discriminate between those persons who purport to have high self-esteem and those with low self-esteem. The original RSE scale appears to have acceptable reliability and validity.

An Alternative Scoring Method

The RSE-40

Although Rosenberg (1965, 1979) used the "contrived" Guttman scaling method, the individual items actually use the classical Likert response continuum (See Appendix A). The alternative Likert scoring procedure, labeled the **RSE-40** for lack of a better acronym, simply adds the typical Likert metric scaling system to the RSE response continuum. In other words, each scale item receives a metric score from one (1) to four (4). These individual item scores are then summed, across items, to provide a total score for each respondent. The RSE-40 total score may range from a low of 10 to a high of 40. This scaling method is referred to as a "summated rating scale" and meets the conditions for unidimensional scaling.

Likert Scaling

Likert scales are popular scalars due to their flexibility, ease of construction, and ease of scoring. The original Likert scaling method (Likert, 1939) was developed as a five point scoring system to more easily rate a person's attitudinal response toward a specific set of statements (items). The five response categories are: strongly agree (SA), agree (A), undecided (U), disagree (D),

and strongly disagree (SD). However, in practice, the undecided category is often dropped so that there is a four response category: (SA), (A), (D), and (SD). Likert scaling procedures simply add metric weights to each of the response categories. Here are two examples of how the weights are applied, using one positive and one negative statement from the 1979 RSE. Note that in example A (positive statement) **strongly agree** carries the four (4) weight and that in example B (negative statement) **strongly disagree** carries the four (4) weight. Since it is typical for the Likert method to have both positive and negative statements special attention should be

---

Example A (Positive):

Item (1) On the whole, I am satisfied with myself.

Response Category ------>	SA	A	D	SD
Metric ------>	(4)	(3)	(2)	(1)

Example B (Negative):

Item (2) At times I think I am no good at all.

Response Category ------>	SA	A	D	SD
Metric ------>	(1)	(2)	(3)	(4)

---

be given to the reverse scoring in negatively stated items. The Likert system creates a **total scale score** for each individual by summing all the item scores.

RSE-40 Literature Review

The Likert or "summated rating method" of scoring the RSE has been periodically used in the research literature since 1969 (Byrne, 1983; Byrne & Shavelson, 1987; Crandall, 1973; Ellis & Taylor, 1983; Friedel, Hansen, Hummel & Shaffer, 1979; Hensley & Roberts, 1976;

Hunt & Hardt, 1969; Kahle, 1976; Kaplan, 1970; Kaplan & Pokorny, 1969, 1971; Lee, 1972; Miller, 1973; Reynolds, et al., 1980; Reynolds, In Press; Weiss, 1977; Weiss & Knight, 1980). The use of this method of scoring and the reported psychometric characteristics of the RSE-40 are presented in summary form in Table 1. There were 16 studies located which specifically reported using the 40 point Likert scoring system and two (Kaplan & Pokorny, 1969, 1971) inferred its use in their factor analytic study.

Table 1 about here

Descriptive Statistics. There were reports from three types of population samples; high school, college, and adult groups. The sample sizes ranged from 41 to a high of 929. Sample means ranged from 26.25 to 33.31 and standard deviations ranged from 3.45 to 4.8. One study reported descriptive statistics by race and two studies reported them by gender.

Reliability and Validity. Eight studies reported coefficients of reliability; one each for true-score variance and test-retest, two for split half, and four for coefficient alpha. These coefficients ranged from .61 to a high of .87. Eleven of the studies reported validation data using the RSE-40 with other self-inventories, social desirability, locus of control, and several task specific variables. The validity coefficients ranged from -.42 to .75. Two studies reported factor analytic studies for construct

validity and were discussed earlier under the section on factor analysis. These findings concur with those of Ellis & Taylor (1983) who state, "Previous research has shown the scale to be both reliable and valid..." (p. 634).

Conclusions and Summary

Rosenberg (1965, 1979) used the Guttman scaling technique to develop the RSE and the "contrived scaling" method for Guttman scales to score the original RSE. However, besides being difficult to score, there are criticisms of the Guttman scaling technique in the psychometric literature. Both the Guttman scaled RSE and the Likert scaled RSE appear to be unidimensional in make-up and seem to have adequate reliability and validity. However, since the response continuum of the original RSE has all the characteristics of the well regarded Likert scaling method and since the 40 point Likert method has been used in other research studies it is recommended that the Likert scoring system be used. For lack of a better acronym, the label **RSE-40** has been applied to this procedure. By using the RSE-40 summated rating procedure, a wider total score range of 10 to 40 is produced as compared to the more restricted range of zero to six in the Guttman scaled original. This increased range adds variability which provides for better discrimination among individuals. The increased variability should improve the scales use in statistical analyses and allow for more precise psychometric estimates self-esteem.

If one looks at the studies summarized in Table 1, as a group, and is willing to assume future trends, some generalizations may be made. Overall RSE-40 mean scores will tend to range between 26 and

34 with standard deviations between 3.00 and 5.00. Reliability coefficients will probably range between .60 and .87 depending upon the method used. Coefficient alpha, which represents an estimate of internal consistency, should be especially stable ranging between .75 and .85. Based upon these alphas, future factor analytic studies will probably report findings similar to Hensley & Roberts (1976), with any emerging factors representing "response set" rather than unique factors. Other validity studies will probably find the RSE-40 correlating higher with other global measures of self and lower with more specific self measures such as academic self and other task specific measures of self. The scale does not appear to be correlated with social desirability and has little relationship with locus of control. When classified into high/low self-esteem groups the RSE-40 seems to be able to distinguish among several personality and task specific variables.

References

Byrne, B.M. (1983). Investigating measures of self-concept. Measurement and Evaluation in Guidance, 16 (3), 115-126.

Byrne, B.M., & Shavelson, R.J. (1987). Adolescent self-concept: Testing the assumption of equivalent structure across gender. American Educational Research Journal, 24, 365-385.

Crandall, R. (1973). The measurement of self-esteem and related constructs. In J.P. Robinson and P.R. Shaver (Eds), Measures of social psychological attitudes. Ann Arbor, MI: Institute of Social Research, University of Michigan.

Edwards, A.L. (1957). Techniques of attitude scale construction. NY: Appleton, Century & Crofts.

Ellis, R.A., & Taylor, M.S. (1983). Role of self-esteem within the job search process. Journal of Applied Psychology, 68 (4), 632-640.

Fischer, E.H. (1972). Birth planning of youth: concern about overpopulation and intention to limit family size. American Psychologist, 27, 951-958.

Friedel, K., Hansen, J., Hummel, T.J., & Shaffer, W.F. (1979). Self-esteem and SCII: Empirical investigation of a clinical observation. Measurement and Evaluation in Guidance, 12 (2), 82-86.

Hensley, W.E., & Roberts, M.K. (1976). Dimensions of Rosenberg's Self-Esteem Scale. Psychological Reports, 38, 583-584.

Hunt, D.E., & Hardt, P.H. (1969). The effect of Upward Bound programs on the attitudes, motivation and academic achievement of Negro students. Journal of Social Issues, 25(3), 117-129.

Kahle, L.R. (1976). Comparison of four methods of measuring self-esteem. Psychological Reports, 39, 974.

Kaplan, H.B. (1970). Recent life experiences, Archives of General Psychiatry, 22, 324-331.

Kaplan, H.B., & Pokorney, A.D. (1969). Self-derogation and psychosocial adjustment. Journal of Nervous and Mental Disease, 149, 421-434.

Kaplan, H.B., & Pokorney, A.D. (1971). Self-derogation and childhood broken homes. Journal of Marriage and Family, 33, 328-337.

Kohn, M.L. (1969). Class and conformity: A study in values. Homewood, IL: Dorsey.

Lee, R.E. (1972). A note on the relationship of critical behavior and self-esteem in the hospitalized veteran. Psychotherapy: Theory, Research and Practice, 9, 421-434.

Lieberman, M.A., Yalim, I.D., & Miles, M.B. (1972). The impact of encounter groups on participants: Some preliminary findings. Journal of Applied Behavioral Studies, 8, 29-50.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140.

Miller, T.W. (1973). Male self-esteem and attitudes toward women's roles. Journal of College Student Personnel, 14, 402-406.

Nocks, J.J., & Bradley, D.L. (1969). Self-esteem in an alcoholic population. Diseases of the Nervous System, 30, 611-617.

Nunnally, J. (1975). Psychometric theory (2nd ed.). NY: McGraw-Hill.

Reynolds, W.M., Ramirez, M.P., Magrina, A., & Allen, J.E. (1980). Initial development and validation of the Academic Self-Concept Scale. Educational and Psychological Measurement, 40, 1013-1016.

Reynolds, W.M. (1988). Measurement of academic self-concept in college students, Journal of Personality Assessment, 52(2), 223-240.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). Conceiving the self. NY: Basic Books.

Silber, E., & Tippett, L. (1965). Self esteem: Clinical assessment and measurement validation. Psychological Reports, 16, 1017.

Stang, D.J. (1972). Conformity, ability, and self-esteem. Representative Research in Social Psychology, 3, 97-103.

Stouffer, S.A., Borgatta, E.F., Hays, D.G., & Henry, A.F. (1952). A technique for improving cumulative scales. Public Opinion Quarterly, 16, 273-291.

Tessler, R.C., & Schwartz, S.H. (1972). Help seeking self-esteem, and achievement motivation: An attributional analysis. Journal of Personality and Social Psychology, 21, 318-326.

Torgerson, W. (1958). Theory and methods of scaling. NY: Wiley.

Weiss, H.J. (1977). Subordinate imitation of supervisor behavior: The role of modeling in organizational socialization. Organizational Behavior and Human Performance, 19, 89-105.

Weiss, H.J., & Knight, P.A. (1980). The utility of humility: Self-esteem, information search and problem solving efficiency. Organizational Behavior and Human Performance, 25, 216-223.

Williams, J.A., & Stockton, R. (1972). Black family structures and functions: An empirical examination of some suggestions made by Billingsley. Journal of Marriage and Family, 35, 39-49.

Wylie, R. (1974). The self-concept: Vol. 1 (rev. ed.). Lincoln, NE.: University of Nebraska Press.

Yamamoto, K., & Wiersma, J. (1967). Rejection of self and other deviant others among student teachers. Journal of Special Education, 1 (4), 401-407.

Appendix A

(1)	On the whole, I am satisfied with myself.	SA (4)	A (3)	D (2)	SD (1)
*(2)	At times I think I am no good at all.	SA (1)	A (2)	D (3)	SD (4)
(3)	I feel that I have a number of good qualities.	SA (4)	A (3)	D (2)	SD (1)
(4)	I am able to do things as well as most other people.	SA (4)	A (3)	D (2)	SD (1)
*(5)	I feel I do not have much to be proud of.	SA (1)	A (2)	D (3)	SD (4)
*(6)	I certainly feel useless at times.	SA (1)	A (2)	D (3)	SD (4)
(7)	I feel that I'm a person of worth, at least on an equal plane with others.	SA (4)	A (3)	D (2)	SD (1)
*(8)	I wish I could have more respect for myself.	SA (1)	A (2)	D (3)	SD (4)
*(9)	All in all, I am inclined to feel that I am a failure.	SA (1)	A (2)	D (3)	SD (4)
(10)	I take a positive attitude toward myself.	SA (4)	A (3)	D (2)	SD (1)

---

* Indicates negatively stated items with reverse scoring.

Rosenberg Self-Esteem Scale (RSE) (Rosenberg, 1979).

Table 1: Summary of Research Reports using the Forty (40) Point Likert Scoring System for the Rosenberg Self-Esteem Scale.

Source	Type Sample	N	Statistics Mean	S.D.	Reliability Type	Coefficient	Validity Coefficients	RSE with
Byrne & Shavelson (1987)	High School	832			True Score	.74(Male) .87(Female)		
Byrne (1983)	High School	929			Test/retest	.61	.60 .37 .37	Coopersmith SE Coopersmith SE (Academic Subscale) Brookover SC Ability
Crandall (1973)							.60	Coopersmith SE
Ellis & Taylor (1983)	College	86					-.20 to .54	Task Specific SE
Friedel et al., (1979)	College	198	31.15	4.2	Split Half	.74	.21	SCII Part VII
Hensley & Roberts (1976)	College	479						Factor Anaylsis
Hunt & Hardt (1969)	High School	1966 211(Black) 90(White) 1968 211(Black) 90(White)	28.73 28.01 30.27 29.53					
Kahle (1976)	College	442(Total) 194(Male) 248(Female)	31.70 32.40 31.10	4.0 4.1 3.9	Odd/Even	.80	.60 .75	Self Description Inventory Feelings of Inadequacy
Kaplan & Pokorny (1969)	Adults	500						Factor Analysis
Lee (1972)	Adults	53(Male)	Range (20-39)				Difference High/low SE and negativeisa	
Miller (1973)	College	171(Male)	26.25	3.45			-.14 to .49	Women's Liberation Questionnaire
Reynolds (1988)	College	589	31.46	4.8	Alpha	.83	-.21 .10 .04 -.05	Academic SC Social Desirability Rotter IE GPA
Reynolds et al. (1980)	College	427			Alpha	.82	.45 -.01	Academic SC GPA
Weiss & Knight (1980)	College	41(Male)	32.33	3.61	Alpha	.76	-.42 to .41	Task Specific
Weiss (1977)	Adult	141	33.31	3.48	Alpha	.75	Significant SE r differences between high/low SE on success, coapetence, reward power.	

BEST COPY AVAILABLE